Seoul, 23-25 September 2025

1ST INTERNATIONAL SYMPOSIUM ON THE MODERNIZATION OF LABOUR STATISTICS

INTEGRATING LABOUR MARKET SURVEYS WITH STATISTICAL REGISTERS: EXPERIENCES FROM ITALY'S NATIONAL INSTITUTE OF STATISTICS

Federica Pintaldi
Istat | Italy

Summary

- Introduction
- Using educational attainment as an auxiliary variable in the Labour Force Survey
- Employees' gross earnings from the Labour Register in the Labour Force Survey
- Data on the public sector directly from the Labour Register in the Structure of Earnings Survey (SES)
- Future prospects
- Conclusion

Introduction

- In recent years, Istat has been implementing the **Italian Integrated System of Statistical Registers** (ISSR), a harmonized set of statistical registers that integrate different sources of data (Population Register, Labour Register, Business Register, etc.)
- The construction of each Register is based on a **plurality sources (mainly administrative)**, each with different coverage, variables, units, reference periods, metadata, timeliness, etc.
- The Register is produced through a **complex process of harmonizing and integrating** different administrative data, which requires diverse skills (methodological, statistical, subject-matter, and IT)
- These **new sources** open up the possibility of **integrating such data at different stages of the survey process**: for sample design and selection, non-response adjustment, weighting, validation, use as auxiliary variables, imputation, model-based estimation, as well as expanding the set of available information through record linkage

Register variables as auxiliaries for the LFS

Istat

Education level as auxiliary variable

- In recent years, due to the **increasing difficulties in maintaining high response rates** and ensuring good representativeness of collected samples, quality issues in surveys have become more frequent
- According to **European Regulations these control totals applied in calibration** derive from demographic administrative sources. This auxiliary information is incorporated in the estimation procedure with the calibration estimator to improve the quality of the estimates
- **Population Registers** provide information on sex, age, citizenship, household structure, and education level of the resident population. This information plays an important role **both in the treatment of total non-response and in the final calibration**
- The **LFS weighting procedure has three steps**: (1) initial weights are calculated as the inverse of the inclusion probabilities; (2) intermediate weights are obtained by applying non-response corrections based on age, gender, citizenship, household size and household types); (3) final weights are derived by solving a constrained minimization problem through a calibration model, ensuring consistency with population totals and equal weights for all household members

Education level as auxiliary variable

- On **January 2021 new population** figures were made available for the period 2011-2021, according to the results of the Population continuous Census 2018. Consequently, LFS weights were recalculated to be coherent with Census population

- The availability to **link the actual sample to the Population Register data made it possible to review the non-response treatment**. Correction factors for non-response treatment are calculated with reference to the estimates obtained from LFS theoretical sample, through an iterative proportional fitting procedure

- Given the important **relationship between education level and the main labour market indicators**, an experimental study was carried out to assess the effect of integrating this information, as an auxiliary variable, from the Population Register into the LFS estimates

- The **use of education as auxiliary information in the LFS calibration reduced the bias** due to the unbalanced coverage of certain population subgroups, and was therefore **adopted in the new estimation methodology**

- **Further studies** will be useful to refine the model, but this work has already shown that the integration of administrative sources throughout statistical registers and their use in the calibration model represent a **promising way to improve the quality of LFS estimates**

Gross monthly pay for LFS employees

Gross monthly pay in the LFS

- Following the adoption of the **IESS framework regulation (Reg. 1700/2019)**, starting from 2021 the LFS must provide a variable on **gross monthly pay for employees** (INCGROSS), which replaces the INCOME variable collected till 2020 on net monthly pay

- In Italy, directly **asking interviewees for their gross salary is not feasible**, as they may not be aware of the gross amount (employees receive net salary, after taxes and contributions)

- On the other hand, **administrative information can be used** to this aim through Labour Register (LR) which contains extensive information on employment, including wages

- The new variable INCGROSS is obtained adopting a **mixed strategy** based on register and survey data, and also model prediction:
  - ✓ LFS collected data are **linked with the LR** and, subject to a validation step, information on the wage is imputed from the LR;
  - ✓ **Model** prediction is applied in cases where LR **information is not validated** (inconsistent work characteristics between LFS and LR, outliers) **or not available** (missing value or missing linkage)

Gross monthly pay in the LFS

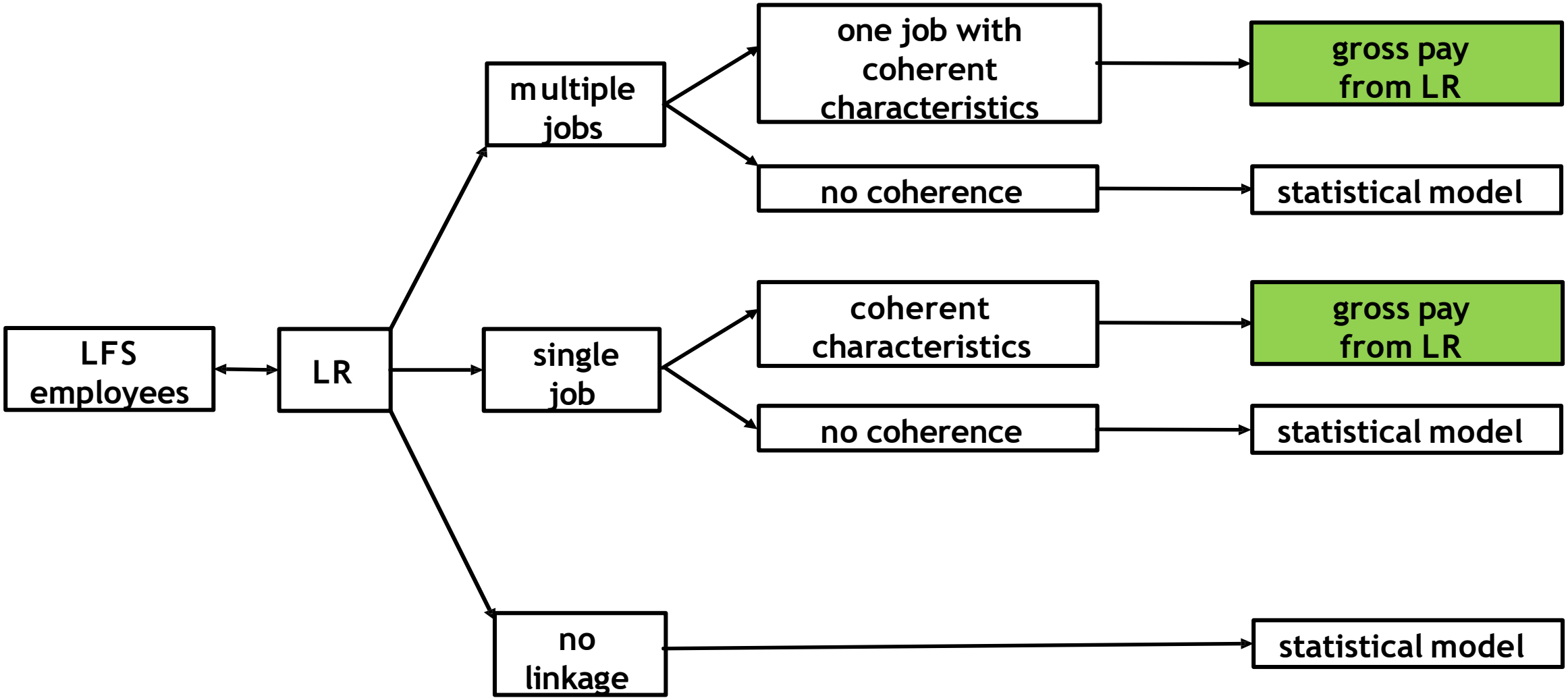
The **information available from LR has been partially processed** to derive the variable responding to the specific definition requirements of regulations of LFS:

- ✓ additional monthly payments (e.g., annual bonuses such as the 13th month) have to be allocated proportionally to the monthly wage received
- ✓ in public sector, unlike the private sector, maternity and parental leave are included in declared gross wages and cannot be separately and clearly identified as social benefits. To harmonize the estimate of gross wage between private and public sector, the allowances payable by social security institutions were added to the gross wage also in the private sector

The **validation process** consists of two steps:

1. **Each LFS employee was linked with one or more LR job positions active during the reference week** (using the unique pseudo-anonymized person ID that complies with privacy requirements)
2. **Consistency between LFS and LR job characteristics** was validated using three variables: full time/part time, temporary/permanent, economic activity (NACE). Working time is considered the most relevant variable, as it determines different pay levels

Gross monthly pay in the LFS



Gross monthly pay in the LFS

To develop the statistical models aimed at predicting gross wages in cases where it is not possible to impute from the LR, **three groups** have been identified:

1. LFS employees not **linked with the LR or linked but not validated** (excluding irregular employment):
   - ✓ Gross wage information in the LR is missing or anomalous (outlier)
   - ✓ LFS part-time workers linked but full-time in the LR

2. LFS full-time workers linked with the LR but **partially validated** because recorded as part-time in the LR (public sector excluded): the presence of **'grey' work** must be verified (partially regular employment, where part of the wage is undeclared in administrative sources)

3. LFS employees not linked with the LR or linked but not validated, where **irregular employment is presumed**; a comparison has also been made with the official estimates of National Accounts

The integration of survey and register data has **enriched the information** available and, at the same time, to **capture new phenomena** that cannot be identified by a single source (not possible from registers and only partially from surveys)

Istat

Structure of Earnings Survey (SES)

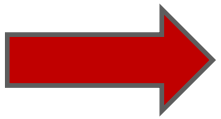
Istat

Structure of Earnings Survey (SES)

- The **Structure of Earnings Survey (SES)** is conducted in EU countries, candidate countries, and European Free Trade Association (EFTA) countries

- It is a **large-scale enterprise survey** on the relationship between pay levels and individual employee characteristics (sex, age, occupation, length of service, highest educational level attained, etc.), as well as employer characteristics (economic activity, enterprise size, and location)

- The statistics cover enterprises with at least **10 employees operating in all areas of the economy, except public administration**. Some countries also provide information on a **voluntary basis** on public administration



Italy is among the countries that **also provide information on public administration** on a voluntary basis **using the Labour Register (LR)**

Istat

Structure of Earnings Survey (SES) and administrative data

The process for the private sector

- At the core of the statistical process for the private sector lies the Labour Register (LR), which is used in all phases of the survey: sampling with a reduction of the sample size, direct sample of employees, simplification of the questionnaire by eliminating variables already available from the register and pre-filling questionnaire with LR information (mainly on wages and hours), control and correction, where register data have been used as parameters in the reconstruction of non-responses, calibration to register known totals

The process for the public sector

- The set of economic units in the public sector derives from the Public Institutions Register, which consists of all institutional units of the General Government sector of the European System of Accounts, as well as public institutions. The target population of employees was drawn from those LR units belonging to the Public Institutions Register. From this population, a sample was selected using a design comparable to that of the private sector. On this sample, the SES variables of interest were populated exclusively using the LR

The use of a mixed register–survey process has **reduced respondent burden** (with some information taken directly from registers and partly pre-filled questionnaire), **expanded the available information**, and **improved data quality**

Future prospects

Comparing household survey data with administrative data

- **Observation field** of the source
- **Time period** covered (e.g. occasional work compared with the reference week in LFS)
- Distinction between **de facto and contractual** employment status (e.g. part-time contract with actual full-time hours). Survey required in these cases

Errors present in all sources
Distinguishing between **data inconsistencies** and **different meanings** of phenomena

The integration of household survey microdata with registers improves the understanding of phenomena, reducing weaknesses and enhancing the strengths of the different sources, and makes it possible to study phenomena that cannot be identified when using a single source

Istat

Labour Force Survey compared with Register data

Labour force survey Strengths	Register data
Average stock of employed persons over a given period of time	Not always possible
All forms of regular employment across all sectors of economic activity	Potentially Strengths
Irregular employment (fully or partially)	Weaknesses, but used for validation
Quality of work and subjective indicators	Weaknesses
Employment status of the entire population aged 15 and over	Weaknesses
Typology of the unemployed and the inactive population	Weaknesses
Data timeliness	Weaknesses

Labour force survey Weaknesses	Register data
Population not residing in Italy or not living in households	Strengths
Sampling errors	Strengths
Estimation of rare phenomena	Strengths
Employment histories over time	Strengths
Granular territorial level	Strengths
Enterprise-specific characteristics	Strengths
Type of employment contract	Strengths

Conclusion

- **Micro-level integration** of sources represents a **new phase** in the use of statistical data, allowing a more accurate reading of complex labour market dynamics
- Careful analysis is needed to **distinguish between errors, false declarations, and differences in the meaning** of phenomena
- The integrated use of survey and register data **strengthens the reliability of analyses**, as their complementarities compensate for individual weaknesses
- This integration not only increases accuracy but also **expands the information base** (e.g., longitudinal data to reconstruct work histories, school-to-work transitions, household income, etc.)
- Enabling the study of new and **hidden labour market phenomena** offers powerful tools for deeper insights and innovative labour market research on employment and labour market trends (e.g. false part-time work, undeclared work, etc.)

Thank you

감사합니다

Grazie

Federica Pintaldi | pintaldi@istat.it

Istat | Istituto Nazionale di Statistica

References

Iorio, A., & Martini, A. (2025). *Possible improvements in the estimation of IT-LFS through integration with the Registers System*. Rivista Italiana di Economia Demografia e Statistica (RIEDS), Jan–Mar 2025.
https://www.rieds-journal.org/rieds/article/view/290

Ceccato, F., Ciarallo, M. A., & Conigliaro, P. (2022). *Exploring earnings differences hidden behind the Gender Pay Gap*. Rivista Italiana di Economia Demografia e Statistica (RIEDS), 76(1), 68–78
https://www.rieds-journal.org/rieds/article/view/101/94

Anitori, P., De Gregorio, C., Esposito, L., Fioroni, L., Giordano, A., & Pintaldi, F. (2019). *Income and job trajectories of Italian employees: An analysis of longitudinal indicators obtained through the micro integration of LFS and statistical registers*. Meeting of the Group of Experts on Measuring Quality of Employment, Geneva, Nov 6–8, 2019. UNECE.
https://unece.org/sites/default/files/datastore/fileadmin/DAM/stats/documents/ece/ces/ge.12/2019/S5_Italy_paper.pdf

Baldi, C., Ceccarelli, C., Gigante, S., Pacini, S., & Rossetti, F. (2018). *The Labour Register in Italy: The new heart of the system of labour statistics*. Rivista Italiana di Economia Demografia e Statistica (RIEDS), 72(2), April–June 2018.
https://www.sieds.it/listing/RePEc/journl/2018LXXII_N2_RIEDS_10_64_Baldi_Ceccarelli_Gigante_Pacini_Rossetti_ok.pdf

Rocci, F., Pacini, S., Serbassi, L., Sorrentino, M., & Congia, M. C. (2017). *A new estimate of hours worked based on integration of survey and admin data*. Rivista Italiana di Economia Demografia e Statistica (RIEDS), LXXI(3), July–Sep 2017.
https://www.sieds.it/listing/RePEc/journl/2017LXXI_N3_RIEDS_12_Rocci_Pacini_Serbassi_Sorrentino_Congia.pdf

Rocci, F., Pacini, S., Serbassi, L., Sorrentino, M., & Congia, M. C. (2016). *A prediction approach for the estimation of hours worked using integrated register and survey data*. Rivista di Statistica Ufficiale, 3, 41–71. ISTAT.
https://www.istat.it/wp-content/uploads/2020/04/RSU_3-2016.pdf

Istat